Amino acid intrinsic α -helical propensities III: Positional dependence at several positions of C terminus

MICHAEL PETUKHOV,^{1,2} KOICHI UEGAKI,³ NOBORU YUMOTO,³ AND
LUIS SERRANO¹

¹European Molecular Biology Laboratory (EMBL), Heidelberg, D-69012, Germany

²Division of Molecular and Radiation Biology, St. Petersburg Nuclear Physics Institute, RAS, Gatchina, 188350, St. Petersburg, Russia

³Osaka National Research Institute, AIST, Ikeda, Osaka 563, Japan

(RECEIVED June 27, 2001; FINAL REVISION December 7, 2001; ACCEPTED December 7, 2001)

Abstract

In this study, we have analyzed experimentally the helical intrinsic propensities of noncharged and non-aromatic residues at different C-terminal positions (C1, C2, C3) of an Ala-based peptide. The effect was found to be complex, resulting in extra stabilization or destabilization, depending on guest amino acid and position under consideration. Polar (Ser, Thr, Cys, Asn, and Gln) amino acids and Gly were found to have significantly larger helical propensities at several C-terminal positions compared with the α -helix center (-1.0 kcal/mole in some cases). Some of the nonpolar residues, especially β -branched ones (Val and Ile) are significantly more favorable at position C3 (-0.3 to -0.4 kcal/mole), although having minor differences at other C-terminal positions compared with the α -helix center. Leu has moderate (-0.1 to -0.2 kcal/mole) stabilization effects at position C2 and C3, whereas being relatively neutral at C1. Finally, Met was found to be unfavorable at C1 and C2 ($+0.2$ kcal/mole) and favorable at C3 (-0.2 kcal/mole). Thus, significant differences found between the intrinsic helical propensities at the C-terminal positions and those in the α -helix center must be accounted for in helix/coil transition theories and in protein design.

Keywords: C terminus; secondary structure; folding; α -helix

The intrinsic secondary structure propensities are defined as the free energy cost required to fix an amino acid in helical angles, excluding the main-chain–main-chain hydrogen contribution, side-chain–side-chain interactions and electrostatic interactions with the helix macrodipole (Muñoz and Serrano 1995a). These propensities have been the subject of many experimental (Chakrabarty et al. 1991, 1994; Qian 1993) and theoretical studies (Muñoz and Serrano 1994b; Swindells et al. 1995; Stapley and Doig 1997). Intrinsic helical propensities of natural amino acids were also considered to be an important factor affecting protein stability

(Villegas et al. 1995; Petukhov et al. 1997; Facchiano et al. 1998; Strop et al. 2000) and kinetics of protein folding (Kiefhaber and Baldwin 1995; Taddei et al. 2000). Secondary structure propensities of different amino acids have been interpreted in terms of van-der-Waals' (Street and Mayo 1999) and electrostatic interactions (Avbelj and Moulton 1995) between amino acid side chains and peptide backbone, losses of configurational entropy upon folding (Creamer and Rose 1994; Lee et al. 1994), hydrophobic effects (Blaber et al. 1993), and a combination of the above-mentioned contributions to free energy (Petukhov et al. 1998).

The assumption of position independence of α -helix propensities was made in many statistical-mechanical models describing the helix-coil transitions in peptides (Finkelstein et al. 1991; Doig et al. 1994; Muñoz and Serrano 1994a).

Reprint requests to: Dr. Luis Serrano, European Molecular Biology Laboratory (EMBL), Meyerhofstrasse 1, Heidelberg D-69117, Germany; e-mail: serrano@EMBL-Heidelberg.de; fax: 49-6221-387-306.

Article and publication are at <http://www.proteinscience.org/cgi/doi/10.1110/ps.2610102>.

However, recent experimental and theoretical studies showed that this assumption is not valid, for at least some natural amino acids in the first (N-terminal) turn of an α -helix (Lacroix et al. 1998; Petukhov et al. 1998, 1999; Sun et al. 2000; Cochran and Doig 2001; Cochran et al. 2001). This non-ideality for helix secondary structure propensities was first introduced into the helix/coil transition algorithm AGADIR1s-2 (Lacroix et al. 1998). The positions in the first and the last α -helix turns are not equivalent to positions in the center of α -helix. Amino acid side chains at the helix termini are more solvent exposed and have a smaller number of intramolecular Van-der-Waals' contacts. In an ideal α -helix, due to its Christmas tree-like arrangement of amino acid side chains, the entropy of residues in the center and near the C-terminal positions is approximately equal, whereas this is not the case for the N-terminal turn (Lacroix et al. 1998; Petukhov et al. 1998, 1999; Sun et al. 2000; Cochran and Doig 2001; Cochran et al. 2001). However, in many protein α -helices, ϕ and ψ dihedral angles at positions C1 and C2 are known to be distorted (Strehlow and Baldwin 1989; Chakrabartty et al. 1993), allowing some extra flexibility of the side chains (Lacroix et al. 1998).

In the two previous studies of these series, we have examined the positional dependence of amino acid α -helical propensities at N-terminal positions (Petukhov et al. 1998, 1999). The effect was found to have a complex nature and varies in magnitude and sign for different amino acids. More recently, Thomas et al. (2001) analyzed the propensities of the amino acids to be at position C' of the C-capping box of α -helices. In this study, we have examined the position dependence at the C terminus of α -helices. To do so, we have synthesized and analyzed, by far-UV circular dichroism (CD), a series of 18 residue Ala-based peptides substituted with Gly, Ser, Thr, Asn, Gln, Val, Ile, Leu, Met, and Cys at C1, C2, C3, and C7 positions (Richardson and Richardson 1988). These template peptides showed no aggregation complications and allow measuring without influence of side-chain-side-chain interactions, the helical propensities at the C-terminal and central positions of an α -helix. The results of CD measurements were interpreted in free-energy terms by use of the helix/coil transition theory [AGADIR1s, (Muñoz and Serrano 1995a,b)], with some modifications to include position dependence of propensities [AGADIR1s-2, (Lacroix et al. 1998)].

Results

The helical intrinsic secondary structure propensities were defined (Muñoz and Serrano 1995a) as the free energy required to fix an amino acid in helical angles, excluding the main-chain-main-chain hydrogen contribution, local interactions with other amino acids in the peptide chain, and electrostatic interactions with the helix macrodipole. Thus,

all of the factors related to flanking residues or with the helical structure of the peptides are excluded from the definition. Particularly, the definition of amino acid intrinsic helical propensities does not include main-chain-main-chain hydrogen bonding. A similar approach in which amino acid independent energy terms are separated from intrinsic helical propensities have been used by other groups as well (Creamer and Rose 1994; Lee et al. 1994; Avbelj and Moulton 1995; Street and Mayo 1999). Another version of statistical mechanical theory used widely in the field, the Lifson-Roig theory, includes the hydrogen bond contribution into the intrinsic propensity (Stapley et al. 1995). Nevertheless, as has been shown, AGADIR and the Lifson-Roig theory produce virtually the same numbers for the elongation and nucleation parameters when fitted to the same set of peptides. (Muñoz and Serrano 1997).

CD measurements

In this work, we synthesized a series of 52 peptides using a newly designed 18-residues host peptide:

The N terminus was always free (Fr), and the C terminus was free or amidated (Fr/Am). N-terminal Tyr, separated by two Gly residues, has been used for accurate measurements of peptide concentration. The Gly spacer is there to diminish the aromatic contribution (Chakrabartty et al. 1993). There is also a Capping Box sequence (Harper and Rose 1993) and a possible salt bridge at the N-terminal part of the sequence to stabilize the peptide helix. Lys and Arg residues were added to favor peptide solubility. Positions C1, C2, C3, and C7 have been used to place guest residues as follows: Gly, Ser, Thr, Asn, Gln, Val, Ile, Leu, Met, and Cys.

Figure 1 shows typical CD spectra of all peptides used in this study. Table 1 shows the average mean residue ellipticities and estimations of helical content derived from the mean residue ellipticity as described by Chen et al. (1974). The positional effect can be appreciated from these data even without any precise calculations on the basis of statistical mechanics. Generally, in the absence of any other complications from side-chain-side-chain, capping motifs, and helix macrodipole interactions, we can expect that the helix fraction should monotonously decrease when the position of a guest residue shifts from the peptide C terminus in the direction to its center. This is because the intrinsic propensities of all amino acid residues are worse than that of Ala and, thus, in the absence of side-chain-side-chain interactions, a mutation not involving a charged residue should

Fig. 1. CD spectra of the series of peptides at 10 μ M concentration. The indexes of the peptides and helical content are indicated in the panels. The sequences, mean residue ellipticities, and estimations of helical content are given in Table 1. Other experimental conditions are given in Materials and Methods.

destabilize any helical segment in which it is localized. Given that the number of possible helical segments in which a particular residue can participate in is higher for the central positions than at helix termini, the destabilization effect of centrally disposed residues is, in principle, higher than

that of terminal ones. Thus, any significant deviation from a monotonic decrease of the helical content of peptide helices with centrally disposed guest residues is a direct indication of a positional dependence of amino acid intrinsic helical propensities. We can see from the data present in Table 1

Fig. 1. (Continued.)

that such a deviation does exist for the majority of amino acids used in this study.

Determination of the amino acid helical propensities by use of AGADIR1s-2

In peptide helices, there is not a single α -helix in equilibrium with the coil state, but a broad ensemble of helical

conformations with different lengths starting and ending at different positions of the sequence. Therefore, the experimental helical contents can only be interpreted in energy terms of intrinsic helical propensities by fitting the far-UV CD data to a statistical mechanical model of the helix/coil transition, as described in Materials and Methods. We used the algorithm AGADIR1s-2 (Lacroix et al. 1998) that in-

Table 1. Sequences and results of CD measurements of the helical contents for the series of synthetic peptides used in this study^a

Peptide	Sequence	$-\theta_{222}$ (deg · cm ² /dmole)	Helical content
A1F	YGGSAKEAAARAAAAAA-Fr	-6,822	20.1
A1A	YGGSAKEAAARAAAAAA-Am	-19,063	56.3
T1F	YGGSAKEAAARAAAAAAAT-Fr	-4,367	12.9
T1A	YGGSAKEAAARAAAAAAAT-Am	-12,833	37.9
T2A	YGGSAKEAAARAAAAAATA-Am	-13,544	40.0
T3A	YGGSAKEAAARAAAAAA- <i>Am</i>	-11,529	34.0
T7A	YGGSAKEAAARTAAAAAA- <i>Am</i>	-10,293	30.4
S1F	YGGSAKEAAARAAAAAAS-Fr	-7,144	21.1
S1A	YGGSAKEAAARAAAAAAS-Am	-15,914	47.0
S2A	YGGSAKEAAARAAAAASA-Am	-14,153	41.8
S3A	YGGSAKEAAARAAAAASAA-Am	-13,442	39.7
S7A	YGGSAKEAAARSAAAAAA- <i>Am</i>	-10,124	29.9
N1F	YGGSAKEAAARAAAAAAN-Fr	-5,417	16.0
N1A	YGGSAKEAAARAAAAAAN-Am	-16,252	48.0
N2A	YGGSAKEAAARAAAAANA-Am	-15,135	44.7
N3A	YGGSAKEAAARAAAAANAA-Am	-14,018	41.4
N7A	YGGSAKEAAARNAAAAAAA-Am	-10,598	31.3
Q1F	YGGSAKEAAARAAAAAAQ-Fr	-7,009	20.7
Q1A	YGGSAKEAAARAAAAAAQ-Am	-18,555	54.8
Q2A	YGGSAKEAAARAAAAQA-Am	-17,099	50.5
Q3A	YGGSAKEAAARAAAAQAA-Am	-17,031	50.3
Q7A	YGGSAKEAAARQAAAAAA- <i>Am</i>	-14,593	43.1
G1F	YGGSAKEAAARAAAAAAG-Fr	-6,602	19.5
G1A	YGGSAKEAAARAAAAAAG-Am	-16,252	48.0
G2A	YGGSAKEAAARAAAAAGA-Am	-13,070	38.6
G3A	YGGSAKEAAARAAAAAGAA-Am	-9,396	27.8
G7A	YGGSAKEAAARGAAAAAA- <i>Am</i>	-5,112	15.1
V1F	YGGSAKEAAARAAAAAAV-Fr	-4,842	14.3
V1A	YGGSAKEAAARAAAAAAV-Am	-14,136	41.8
V2A	YGGSAKEAAARAAAAVA-Am	-12,426	36.7
V3A	YGGSAKEAAARAAAAVAA-Am	-14,746	43.6
V7A	YGGSAKEAAARVAAAAAA- <i>Am</i>	-11,072	32.7
I1F	YGGSAKEAAARAAAAAAI-Fr	-4,147	12.3
I1A	YGGSAKEAAARAAAAAAI-Am	-13,950	41.2
I2A	YGGSAKEAAARAAAAAIA-Am	-14,221	42.0
I3A	YGGSAKEAAARAAAAAIAA-Am	-16,371	48.4
I7A	YGGSAKEAAARIAAAAAAA- <i>Am</i>	-13,713	40.5
L1F	YGGSAKEAAARAAAAAAL-Fr	-5,214	15.4
L1A	YGGSAKEAAARAAAAAAL-Am	-16,286	48.1
L2A	YGGSAKEAAARAAAAALA-Am	-17,945	53.0
L3A	YGGSAKEAAARAAAAALAA-Am	-17,675	52.2
L7A	YGGSAKEAAARLAAAAAA- <i>Am</i>	-16,591	49.0
M1F	YGGSAKEAAARAAAAAAM-Fr	-7,957	23.5
M1A	YGGSAKEAAARAAAAAAM-Am	-17,150	50.7
M2A	YGGSAKEAAARAAAAAMA-Am	-15,541	45.9
M3A	YGGSAKEAAARAAAAMAA-Am	-17,624	52.1
M7A	YGGSAKEAAARMAAAAAA- <i>Am</i>	-15,423	45.6
C1F ^b	YGGSAKEAAARAAAAAAC-Fr	-6,772	20.0
C1A ^b	YGGSAKEAAARAAAAAAC-Am	-15,237	45.0
C2A ^b	YGGSAKEAAARAAAAACA-Am	-14,763	43.6
C3A ^b	YGGSAKEAAARAAAAACAA-Am	-14,966	44.2
C7A ^b	YGGSAKEAAARCAAAAAAA- <i>Am</i>	-9,480	28.0

^a Far-UV CD spectra of the peptides were obtained at pH 2 (XF series), or pH 7 for the rest of the peptides, in 5 mM HCl-KCl and Na₂HPO₄-NaH₂PO₄ buffers, respectively, at a temperature of 5°C. Peptide concentrations were ~10 μM and 50 μM. The data of mean residue ellipticity and of helix fraction are an average of CD measurements at 10 μM and 50 μM concentrations. The percentage of α-helix was calculated with the empirical equation $-100 \cdot (\theta_{222} / (39,500(1 - 2.57/n)))$ (Chen et al. 1974); in which *n* is a number of peptide bonds and θ_{222} is an experimentally observed ellipticity of peptide at 222 nm. The error estimates in θ_{222} and in corresponding helical content are based on ~3% errors in peptide concentration measurements.

^b To avoid oxidation of the Cys-containing peptides, 1 mM and 5 mM of β-mercaptoethanol was added to high and low concentrated peptides samples, respectively.

cludes a single-sequence version of the helix/coil transition (Muñoz and Serrano 1997) and a modified set of the free-energy parameters described previously (Muñoz and Serrano 1995a). This is the standard procedure followed by several groups to obtain the N-cap energies (Doig and Baldwin 1995) and the intrinsic helical propensities (Petukhov et al. 1998, 1999) of natural amino acids.

The intrinsic helical propensities obtained in this work are listed in Table 2 (the columns under title $\Delta\Delta G_{\text{exp}}$). Unfortunately, all available statistical mechanical algorithms for description of helix/coil transitions in peptides have a limited accuracy in predictions of experimentally measurable helix fractions of monomeric peptides. This is due to the fact that most of the energy parameters are derived from the experimental data, which also includes errors. Standard errors of AGADIR were reported to be ~7% with an 85% confidence (Lacroix et al. 1998), which approximately correspond to ± 3% absolute error in predictions for peptides having helical content in the range of 40%–50%. Therefore, Table 2 also includes estimates of energy changes corresponding to ± 3% errors in AGADIR predictions for helical contents of peptides under consideration.

A first observation is that, with the exception of Gly, all amino acids are, in general, more favorable at position C3 than in the helix center (Thr within the error). By grouping the amino acids by their chemical properties, one can detect some general behaviors. Hydrophobic residues, Cys, Ile, Val, Leu, and Met are more favorable at position C3 than in the helix center. Regarding positions C1 and C2, polar residues are also more favorable (the only exception being Thr, which has similar values at position C3 and C7). Hydrophobic residues have mixed behaviors at these positions, ranging from being less stable, as in the case of Met at C1 and C2, to being significantly more stable as in the case of Cys. Regarding Gly, we find that it has the largest differences with respect to central positions. Positions C1 and C2 are strongly more favorable (by -1.1 and -0.55 kcal/mole, respectively) compared with the helix center. On the contrary, position C3 was found to be strongly unfavorable by +0.62 kcal/mole.

Statistical survey of the protein database

The amino acid propensities derived from the statistical analysis of a representative set of protein crystal structures was reported earlier to correlate, at least for some amino acids, with the helical propensities measured by CD (Muñoz and Serrano 1994b; Swindells et al. 1995). In our previous study, we also found such a correlation for many natural amino acids at N-terminal positions of α-helix (Petukhov et al. 1998, 1999). Therefore, we expected to find some correlation between C-terminal helical propensities obtained in this study and those found in a statistical survey of protein

Table 2. The positional dependence of the intrinsic helical propensities of natural amino acids at central and three C-terminal positions of the α -helical peptides

Amino acid/ position	ΔE^{ECEPP} (kcal/mole)	ΔE^{hydr} (kcal/mole)	$-T^*\Delta S$ (kcal/mole)	$\Delta\Delta G_{\text{theor}}^{\text{a}}$ relative to position C7 (kcal/mole)	$\Delta\Delta G_{\text{exp}}^{\text{b}}$ relative to position C7 (kcal/mole)	$\Delta\Delta G_{\text{exp}}$ relative to Ala (kcal/mole)
Thr/C1	-29.401	39.226	0.337	0.270	-0.02 ± 0.20	0.80 ± 0.20
Thr/C2	-29.815	39.242	0.268	-0.196	-0.32 ± 0.13	0.50 ± 0.13
Thr/C3	-29.858	39.384	0.324	-0.041	-0.02 ± 0.15	0.80 ± 0.15
Thr/C7	-30.222	39.396	0.717	0.000	0.00 ± 0.12	0.82 ± 0.12
Ser/C1	-28.504	37.288	0.143	0.242	-0.33 ± 0.20	0.52 ± 0.20
Ser/C2	-28.907	37.267	0.102	-0.223	-0.17 ± 0.20	0.68 ± 0.20
Ser/C3	-29.001	37.333	0.126	-0.227	-0.20 ± 0.15	0.65 ± 0.15
Ser/C7	-29.225	37.826	0.084	0.000	0.00 ± 0.13	0.85 ± 0.13
Asn/C1	-31.691	37.014	0.149	0.115	-0.45 ± 0.10	0.35 ± 0.10
Asn/C2	-31.146	36.053	0.180	-0.270	-0.35 ± 0.13	0.45 ± 0.13
Asn/C3	-30.629	35.824	0.146	-0.016	-0.25 ± 0.15	0.55 ± 0.15
Asn/C7	-30.448	35.711	0.094	0.000	0.00 ± 0.11	0.80 ± 0.11
Gln/C1	-35.989	38.895	0.217	-0.027	-0.37 ± 0.24	0.05 ± 0.24
Gln/C2	-35.996	38.801	0.169	-0.176	-0.17 ± 0.15	0.25 ± 0.15
Gln/C3	-36.007	38.857	0.124	-0.176	-0.22 ± 0.15	0.20 ± 0.15
Gln/C7	-35.881	38.821	0.210	0.000	0.00 ± 0.10	0.42 ± 0.10
Gly/C1	-29.615	36.606	—	-0.867	-1.11 ± 0.20	0.42 ± 0.20
Gly/C2	-29.310	36.607	—	-0.561	-0.55 ± 0.25	0.62 ± 0.25
Gly/C3	-29.220	37.040	—	-0.037	-0.62 ± 0.25	2.15 ± 0.25
Gly/C7	-29.227	37.085	—	0.000	0.00 ± 0.20	1.53 ± 0.20
Val/C1	-27.710	38.424	0.281	0.251	-0.18 ± 0.17	0.57 ± 0.17
Val/C2	-28.170	38.573	0.474	0.133	0.00 ± 0.18	0.75 ± 0.18
Val/C3	-28.542	38.506	0.582	-0.198	-0.40 ± 0.13	0.35 ± 0.13
Val/C7	-28.577	38.656	0.665	0.000	0.00 ± 0.10	0.75 ± 0.10
Ile/C1	-27.481	38.137	0.276	0.345	0.05 ± 0.15	0.55 ± 0.15
Ile/C2	-27.864	38.195	0.451	0.195	-0.05 ± 0.15	0.45 ± 0.15
Ile/C3	-28.312	38.127	0.538	-0.234	-0.33 ± 0.12	0.17 ± 0.12
Ile/C7	-28.063	37.984	0.666	0.000	0.00 ± 0.10	0.50 ± 0.10

^a Theoretical difference in free energy with respect to Ala.^b Experimental difference in free energy with respect to Ala.

crystal structures at high resolution. Given the large experimental uncertainties obtained for some of the C-terminal intrinsic helical propensities obtained in this study, it would be difficult to expect a good correlation in all cases. However, for those positions in which significant changes of the intrinsic helical propensities were observed, we could expect some correlation with amino acid frequencies derived from the protein database.

Several groups reported results of statistical analysis of amino acid frequencies at different C-terminal positions in protein α -helices (Aurora and Rose 1998; Kumar and Bansal 1998). However, there is rather a poor correlation ($R = 0.0$ – 0.7 , with an average of 0.33) between amino acid α -helical propensities at positions C1, C2, C3, and central α -helical positions reported by Aurora and Rose (1998) and Kumar and Bansal (1998). Therefore, we re-estimated the amino acid frequencies observed at several C-terminal and central positions of α -helices from a representative set of protein structures at high resolution and $<25\%$ homology as described in Materials and Methods. A total of 1832 α -helices

have been identified in the representative set of protein crystal structures, satisfying the conditions described in the Materials and Methods section. With only a few exceptions (Gly, Cys, and Gln), there is an excellent correlation (0.8 – 0.99) between our data and that reported previously (Kumar and Bansal 1998). As for correlation with data obtained by Aurora and Rose (1998), our results show smaller, although still significant, level correlation with these data ($R = 0.0$ – 0.7 , with an average of 0.42).

Table 3 shows the correlation coefficients calculated for the amino acid propensities of Table 2 and those obtained from the protein database. One can see from the data that significant positive correlation exists only for Ser, Asn, and Met, $R = 0.81$, 0.73 , and 0.97 , respectively. For other amino acids, no correlation or even anticorrelation is observed.

However, it is known that C-termini of protein α -helices are rather quite structurally heterogeneous when compared with center and N-terminal positions. Figure 2 shows the distribution of the main-chain dihedral angles (ϕ , ψ) of all

Table 3. Amino acid frequencies at different positions of protein helices^a

Amino acid	Average of C5–C7	C3	C2	C1	Correlation ^b
THR	61.00	38	65	84	0.11
SER	51.50	52	61	80	0.81
ASN	42.00	57	46	64	0.73
GLN	67.00	70	60	55	−0.62
GLY	30.00	36	26	29	−0.78
VAL	95.50	50	85	51	−0.87
ILE	102.50	67	84	45	−0.05
LEU	157.00	191	133	159	−0.09
MET	39.00	48	28	19	0.97
CYS	11.50	14	11	14	0.39

^a Number of cases found in the protein database.^b Correlation coefficients calculated between the database frequencies and the intrinsic helical propensities obtained from the CD measurements and those listed in Table 2.

natural amino acids at C-terminal positions C1, C2, C3, and at central position CC. One can see from this figure that the heterogeneity of the protein backbone is much higher at C-terminal positions compared with that at central α -helical positions. It is of interest that not only standard deviations of (ϕ , ψ) dihedral angles increase gradually when going from CC to C1 positions as follows: CC (5.8°, 6.10°), C3 (6.1°, 8.9°), C2 (10.2°, 9.8°), C1 (19.5°, 14.7°), but also the mean of the distribution gradually moves from the classical α -helical values of ($\phi \approx -60^\circ$, $\psi \approx -40^\circ$); CC (−63.1°, −42.5°), C3 (−63.6°, −40.5°), C2 (−69.2°, −37.9°), and C1 (−79.1°, −23.9°). Thus, the conformational properties of amino acids at C-terminal positions C1 and C2 significantly deviate from those at other positions within the α -helix, whereas amino acids at positions C3 and CC were found to be more similar structurally.

To take the above into account, similar statistical analysis of protein database was done, assuming the possibility that the backbone dihedral angles of amino acids at C1 position deviate from the classical α -helical angles. Table 4 shows the results of this statistical survey and its correlation with intrinsic helical propensities as obtained by use of CD measurements and statistical mechanics calculations (Table 2). As one can see, allowing more conformational heterogeneity of the amino acid backbone at position C1 drastically improves the correlation with experimental data for the majority of amino acids. This is particularly so for some amino acids (Thr, Ser, Asn, Met, and Gly) in which the correlation is as high as $R \sim 0.8$ – 0.9 . There is also a relatively moderate correlation for Leu ($R = 0.66$). For other residues, the correlation with experimental data is still insignificant (Cys and Ile), or even negative (Val and Gln). Also, it is of interest that correlation between our experimental scale of amino acid helical propensities and amino acid frequencies derived from Protein Database (Table 3) is

relatively high at positions C2 (0.69), C3 (0.49), and central position CC (0.62). The correlation at position C1 is rather weak (0.23).

Calculations based on molecular mechanics

To investigate the role that different physical contributions to free energy may play in the differences of α -helical propensities at C-terminal positions, similar to our previous works (Petukhov et al. 1998, 1999), we performed molecular mechanics calculations for a model peptide α -helix. A minor modification allowing extra flexibility of peptides backbone at position C1 and C2 has been made in the calculation protocol (see Materials and Methods). Table 2 shows the results of the calculations along with the experimental propensities as measured with CD. Although a relatively moderate correlation is observed between theoretical predictions and experimental results ($R \approx 0.6$), in the majority of cases, the magnitude and sign of changes in amino acid α -helix propensities are predicted correctly. Theoretical calculations strongly underestimate the α -helix propensity of polar amino acids (Thr, Ser, Asn, and Gln) at position C1. This can be due to formation of a specific structure allowing the side chains of these residues to make H-bonds with the peptide backbone in a nonstandard helical conformation. The good agreement found between theoretical and experimental data for nonpolar residues at position C1 supports this possibility.

According to both theoretical calculations and experimental measurements (Table 2) polar residues (Thr, Ser, Asn, Gln, and Cys) are generally more favorable at C-terminal positions. This is due mainly to differences in the energy balance [better solvation of polar side chains at the last C-terminal α -helix turn (Thr, Ser, and Cys), smaller nonbonded interactions in the folded state compared with the denatured state (Asn and Gly) and higher configurational entropy of side chains (Thr)], between the C-terminal and central positions of an α -helix.

Nonpolar residues also generally are more favorable in the last helix turn. However, the effect depends on amino acids and positions under consideration and can reverse sign at a neighboring position. Particularly, as shown in Table 2, position C3 is the most favorable one for β -branched residues Ile and Val. Our calculations show that this β -branched residue has only one side-chain conformation ($\chi_1 \approx 180^\circ$) in the helix center and in positions C2 and C3, whereas two conformations are well populated at positions C1 ($\chi_1 \approx 180^\circ$ and -60°). Thus, the higher α -helix propensity of these residues at position C3 is mainly due to entropic reasons. For the other hydrophobic residues, higher configurational entropy still plays a significant role at positions C1, C2, and C3, however, this is largely compensated for by a decrease in the number of van-der-Waals' contacts and

Fig. 2. Distribution of main chain dihedral angles (ϕ, ψ) of all natural amino acids at C-terminal positions C1, C2, C3, and at central position (CC) of protein α -helices derived from a statistical survey of the representative set of protein crystal structures at high resolution (Vriend 1990).

higher exposition to solvent of nonpolar groups. Thus, for Leu and Met, the positional effect is much smaller and, according to both experimental data and theoretical calculation, is within errors.

Similar to our previous study at N terminus of α -helix (Petukhov et al. 1998), Gly showed the highest positional effect among tested amino acids. According to our calculations, this effect can be rationalized in terms of improved solvation of the free carbonyl oxygens of the peptide backbone and lower nonbonded interactions at the C-terminal positions. However, at position C3, the effect of improved solvation is largely canceled out by the hydrogen bond between terminal amide groups, and carbonyl oxygen and nonbonded interactions are similar.

Discussion

Comparison with C-cap values found by other groups

Although the objective of our work was not to redetermine the C-cap contributions of the amino acids considered here, we needed to do it to determine the helical propensities. Having obtained these values, it is interesting to compare them with those obtained by other groups (Doig and Baldwin 1995). In this way, we can find out the importance of specific sequence effects of the particular system used. In the legend to Table 2, we show the C-cap values for: Ala, Gly, Ser, Thr., Asn, Gln, Val, Ile, Leu, and Met. We find that in all cases, except for Ser, the C-cap propensities obtained in

Table 4. Amino acid frequencies allowing conformational heterogeneity at position C1^a

Amino acid	Average of C5–C7	C3	C2	C1	Correlation ^b
THR	29.6	27	51	33	0.97
SER	27.2	31	35	36	0.86
ASN	22.7	32	34	51	0.91
GLN	38.3	41	34	25	−0.69
GLY	20.8	10	12	149	0.77
VAL	46.3	30	26	29	−0.35
ILE	47.7	42	26	26	0.38
LEU	92.0	76	97	53	0.66
MET	25.3	22	10	13	0.77
CYS	4.6	3	8	5	0.03

^a The same as in Table 3. However, amino acids at position C1 were allowed some conformational heterogeneity by using the following WHATIF search motif H/H/H/H/H/H/HT/STC.

^b Correlation coefficients calculated between the database frequencies and the intrinsic helical propensities obtained from the CD measurements and those listed in Table 2.

our work are very similar, or within the reported errors, to those described previously. There is a reasonably good correlation for the two sets of C-cap propensities ($R = 0.6$ and 0.85 with and without Ser, respectively). We also found the same order of C-cap propensities Val < Leu < Asn < Gln < Met as that found by Doig and Baldwin (1995). Given the inaccuracy of the measurements of C-cap propensities due to the relative low contribution of mutations at the C-cap position to the overall helix content of peptides, the correlation between the two data sets is remarkably high.

Local C-cap motifs

One of the problems that could complicate the determination of the helical propensities is the existence of local motifs at the C terminus of α -helices (Aurora and Rose 1998). In AGADIR, two of these motifs, the Schellman (Viguera and Serrano 1995) and the Pro-box (Prieto and Serrano 1997) are considered and parameterized (Lacroix et al. 1998). Thus, when calculating the intrinsic propensity of the amino acids tested here, the possible formation of these motifs and their energy contribution is taken into account. Regarding other motifs, our template peptides were designed to exclude or minimize alternative conformations. However, in some cases, we cannot discard specific nonhelical conformations at the C-cap. Those, if present, could have some impact on the values determined for the helical propensities. However, the fact that we could see similar trends for groups of amino acids having similar chemical properties at the same position suggests that local specific capping motifs do not contribute significantly in our system.

C-cap heterogeneity in α -helices

As we mentioned in the Results section, the C terminus of protein α -helices is quite heterogeneous. In many cases, the

standard α -helical dihedral angles are not found at position C1, in which the ϕ value tends to be larger and the ψ value is smaller. This conformational heterogeneity results in residue C1 being considered to be H or T in DSSP classification. As a result, it is difficult to do a statistical analysis of the protein database, and that is the reason why we improve the correlation with the experimental data when we defined the C-cap position as the first residue following a H or T conformation that is nonhelical. Even after this correction, we still have very poor correlation for position C1, whereas for positions C2 and C3, we find a significant correlation. This could mean that the conformation adopted by our peptides at position C1 is different on average than that found in proteins, in which packing constraints would bias the structure. However, we cannot rule out that local sequence motifs present in proteins will result in a different sequence bias than what should have been obtained if only the secondary structure propensities played an energetic role.

Conclusions

The intrinsic helical propensity of nonpolar and polar amino acids (excluding aromatics, charged residues, and Pro), at C-terminal positions of an α -helix have been analyzed here. The results show a strong dependence of the amino acid propensities on position in the α -helix. The effect depends on the type of amino acid and can be as high 1.1 kcal/mole for Gly at position C1. The magnitude of the effect and its energy significance at several C-terminal positions require accounting for not only in helix/coil transition theory but also in protein engineering.

Materials and methods

Peptide synthesis

The peptides were synthesized on an automated solid-phase peptide synthesizer (Shimadzu PSSM-8), by use of two different kinds of resin (Tenta TGS-RAM for the peptides with amidated C-termini, and Tenta Gel TGS-PHB-Xaa for the peptides with free C-termini) and Fmoc chemistry with benzotriazole-1-yl-oxy-trispyrrolidino-phosphonium hexafluorophosphate and N-hydroxybenzotriazole as coupling reagents. Peptides were cleaved from the resin by trifluoroacetic acid and purified by reverse-phase HPLC on a C18 column. Fmoc-L-amino acids, reagents for peptide synthesis, and Tenta Gel TGS-RAM and TGS-PHB-Xaa resins were purchased from Shimadzu (Kyoto, Japan). The purity of each peptide was assessed by analytical reverse-phase HPLC on a C18 column. Molecular masses were confirmed by mass spectrometry on a time-of-flight mass spectrometer (Shimadzu/Kratos Kompact MALDI II) with matrix-assisted laser desorption ionization.

CD measurements

The CD measurements were done using a Jasco-710 instrument. Each measurement was repeated at least twice using fresh peptides

solutions prepared the same day. The concentration of peptide was determined by UV absorbance of the N-terminal Tyr using the method of Gill and von Hippel (1989). The error in the concentration determination was around 2%. All of the peptides were found to be soluble, at least up to 500- μ M concentration. CD spectra were recorded at pH 7 (Na_2HPO_4 – NaH_2PO_4 buffer solution) and pH 2 (HCl–KCl buffer). All of the spectra were obtained at a temperature of 278K.

To check for concentration dependence of the CD spectra, two different dilutions of the peptides (10 μ M and 50 μ M), were scanned. No concentration dependence of the CD spectra was found in this concentration range. To prevent oxidation in the Cys containing peptides, 1 mM and 5 mM of β -mercapto-ethanol was added to high and low-concentrated samples, respectively. CD spectra in the range of 190 to 250 nm were obtained by use of the continuous scan option (100 nm/min scan speed), with a one-second response time, taking points every 0.2 nm. For every sample, 20 scans were taken. The helical content of the peptides has been estimated using the mean residue ellipticity at 222 nm (Chen et al. 1974).

Calculations based on statistical mechanics

Because short peptides in water are known to populate many different conformations, each of which contribute to the CD signal, CD measurements can be interpreted in terms of changes of amino acid intrinsic helical propensity only, using a statistical mechanics theory. Similar to our previous studies in this series, in this work, the calculations based on statistical mechanics were made with the one-sequence approximation model (AGADIR1s, Muñoz and Serrano 1997), modified to include the possibility of the residue immediately following an acetyl group, or preceding an amide group, to be helical (Lacroix et al. 1998). To include different helical propensities for C-terminal positions, we modified AGADIR1s-2 in the same way as we did before for the N-terminal positions (Petukhov et al. 1998, 1999). In addition to all of the standard AGADIR free-energy contributions (main-chain–main-chain hydrogen bonding, side-chain–side-chain interactions, C- and N-capping interactions, and interactions of charged groups with the helix macrodipole), we included the possibility of different intrinsic helix propensities for amino acids at terminal helix positions C1, C2, and C3 and central positions CC (C4, C5, etc.).

The procedure used for obtaining the actual values of amino acid-intrinsic helical propensities was described in detail previously (Petukhov et al. 1998, 1999) and includes several steps. To accurately obtain the values of intrinsic helical propensities, it is necessary first to measure the C-capping energy for all amino acids under consideration in the context of our particular template Ala-based peptide and to reproduce the helical content of the control peptides within 1% error. Determining the helical content of the nonblocked peptide series (XF) accomplished this. In nonblocked peptides (XF series), the first and last amino acids cannot be helical. Therefore, we can obtain the C-cap contribution of each residue by looking at the helical content of these peptides, bearing different residues at the C-terminal position. The CD analysis is done at pH 2.0 when no electrostatic interactions of the C-terminal COOH group with the rest of the helix are present. The C-capping propensities of all the amino acids used in this study are listed in Table 2.

Once the C-capping values are fixed, then we can proceed to modify the intrinsic propensities of the polar amino acids starting with the first position in the acetylated peptides to reproduce the experimental values. To determine the error, we assumed that the estimation in the helical content is $\pm 3\%$ and proceeded to change

the intrinsic propensity until we reproduced the experimental helical content in percentage $\pm 3\%$. Although the standard deviation between the helical content measured at 10 μ M and 50 μ M is 2.1%, the 3% error is consistent with our previous studies of these series. The procedure includes several steps, when only one energy parameter is allowed to vary at a time as described in Petukhov et al. (1998). A similar approach also has been used by other groups to obtain the free-energy contribution for naturally occurring amino acids at N-cap positions (Doig and Baldwin 1995). In the calculations, the standard AGADIR (Muñoz and Serrano 1997; Lacroix et al. 1998) free energy of hydrogen bond formation in the peptide backbone of -0.895 kcal/mole and Ala intrinsic helical propensity of $+0.618$ kcal/mole were used for all positions of α -helix.

Calculations based on molecular mechanics

The molecular mechanics calculations were used to elucidate energy contributions to intrinsic amino acid α -helical propensities at C-terminal positions C1, C2, and C3 and central position C7 of model 14 residues α -helix having the following sequence:

The calculations were performed as described in Petukhov et al. (1998, 1999) with minor modifications. Energy profiles for the side chains of amino acids under consideration were calculated in the folded as well as in unfolded states for positions C1, C2, C3, and C7 position of the standard helical nomenclature by Richardson and Richardson (1988). N- and C-termini of the model α -helix were free and amidated, respectively. The energy calculations were made with the BKS molecular modeling program (Abagyan and Mazur 1989; Mazur and Abagyan 1989) using the ECEPP/2 force field (Momany et al. 1975; Nemethy et al. 1983). All atoms in the peptides were treated explicitly. Bond lengths and bond angles were fixed at their standard values during the energy calculations and minimization. Van-der-Waals', electrostatic, hydrogen bond, and torsion potentials were included in the energy calculations. Dielectric constant of bulky water ($\epsilon = 81$) was chosen to model water screening of electrostatic interactions in model peptide (Finkelstein 1977; Warshel and Papazyan 1998). The complete list of nonbonded interactions of the peptide under investigation was used to avoid the necessity of an updating of interaction lists during the energy calculations.

The solvation energy term was modeled by a continuum approximation model (Wesson and Eisenberg 1992) for protein-solvent interactions as described in a previous study of this series (Petukhov et al. 1999). Accessible surface area (ASA) was calculated with a computer program implementing the analytical methods for calculation of ASA and its first derivatives (Richmond 1984). Energy profiles were calculated on the grid of ϕ , ψ , χ_1 , χ_2 , and, where applicable, χ_3 with grid steps of 20° and 50 subsequent steps of energy minimization by the conjugate gradient method. The conformation of first N-terminal amino acid (Ser) of the model peptides was fixed as in a Capping Box motif (Harper and Rose 1993; Petukhov et al. 1996). The dihedral angles ϕ , ψ , and ω of other amino acids of the model peptide in the folded state were initially fixed in standard α -helical values of -60° , -40° , and 180° , respectively. For amino acids at C-terminal positions C1 and C2 of α -helix, extra flexibility in the backbone was allowed in the course of the free-energy calculations as follows: in addition to standard

values (-60° , -40°) for ϕ , ψ dihedral angles, amino acids at C1 and C2 position were varied on a regular grid between -30° and -120° (ϕ) and between -60° and 0° (ψ) with a 20° step. In the unfolded state of the model peptide, the ϕ , ψ , and ω angles of all residues were initially set to 180° and allowed to vary (except for the guest residue) by the conjugate gradient energy minimization algorithm. The $-\text{CH}_3$ groups were considered not to contribute to entropy because of its symmetries, and the conformation of this group was initially fixed with dihedral angles of 60° and allowed to vary during the energy minimization. Other dihedral angles of side chains were varied, as described above.

The configuration entropy of amino side chains, S was calculated from the energy profiles at $T = 278\text{K}$ by a classic Boltzmann-Gibbs approach:

$$S = -R \sum P_i \ln(P_i);$$

in which i is the index of summation over all selected conformational states (see below), R is the gas constant, and P_i is the probability of the state, which can be calculated from the canonical Gibbs distribution:

$$P_i = \frac{\exp(-E_{ij}/RT)}{\sum_j \exp(-E_{ij}/RT)}.$$

Index j indicates the summation over the grid points of the phase space that belong to the selected conformational states and for which the energy of system, E_{ij} has been calculated. The conformational states for peptide backbone and of side chains were separated as described by Zimmerman et al. (1977) and by Lee et al. (1994), respectively.

Survey of the Protein Data Bank

The amino acid frequencies at different helix positions were derived with the WHATIF program from 315 protein crystal structures at better than $2.1\text{\AA}$ resolution, with $<25\%$ homology and with R-factor below 0.21 (Vriend 1990). The crystal structures of the proteins were taken from the Brookhaven Protein Data Bank (Bernstein et al. 1977). The sequence motif of H/H/H/H/H/H/H/H/STC to search protein databank was used. Here, S is strand, T is turn, C is coil, and H is helix in the standard WHATIF nomenclature. The amino acid frequencies and the distribution of dihedral angles in peptide backbone and side chains were obtained with WHATIF software for central position C7 as well as for C-terminal positions C1, C2, and C3.

Acknowledgments

M.P. acknowledges financial support from EC grant BIO4-CT97-2086. The authors thank Dr. Vladimir Rybin (EMBL) for useful advice on precise CD measurements.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

References

Abagyan, R.A. and Mazur, A.K. 1989. New methodology for computer-aided modelling of biomolecular structure and dynamics. 2. Local deformations and cycles. *J. Biomol. Struct. Dyn.* **6**: 833-845.

- Aurora, R. and Rose, G.D. 1998. Helix capping. *Protein Sci.* **7**: 21-38.
- Avbelj, F. and Moulton, J. 1995. Role of electrostatic screening in determining protein main chain conformational preferences. *Biochemistry* **34**: 755-764.
- Bernstein, F.C., Koetzle, T.F., Williams, G.J., Meyer, Jr., E.E., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T., and Tasumi, M. 1977. The Protein Data Bank: A computer-based archival file for macromolecular structures. *J. Mol. Biol.* **112**: 535-542.
- Blaber, M., Zhang, X.J., and Matthews, B.W. 1993. Structural basis of amino acid alpha helix propensity. *Science* **260**: 1637-1640.
- Chakrabarty, A., Kortemme, T., Padmanabhan, S., and Baldwin, R.L. 1993. Aromatic side-chain contribution to far-ultraviolet circular dichroism of helical peptides and its effect on measurement of helix propensities. *Biochemistry* **32**: 5560-5565.
- Chakrabarty, A., Schellman, J.A., and Baldwin, R.L. 1991. Large differences in the helix propensities of alanine and glycine. *Nature* **351**: 586-588.
- Chakrabarty, A., Kortemme, T., and Baldwin, R.L. 1994. Helix propensities of the amino acids measured in alanine-based peptides without helix-stabilizing side-chain interactions. *Protein Sci.* **3**: 843-852.
- Chen, Y.H., Yang, J.T., and Chau, K.H. 1974. Determination of the helix and β form of proteins in aqueous solution by circular dichroism. *Biochemistry* **13**: 3350-3359.
- Cochran, D.A. and Doig, A.J. 2001. Effect of the N2 residue on the stability of the α -helix for all 20 amino acids. *Protein Sci.* **10**: 1305-1311.
- Cochran, D.A., Penel, S., and Doig, A.J. 2001. Effect of the N1 residue on the stability of the α -helix for all 20 amino acids. *Protein Sci.* **10**: 463-470.
- Creamer, T.P. and Rose, G.D. 1994. Alpha-helix-forming propensities in peptides and proteins. *Proteins* **19**: 85-97.
- Doig, A.J. and Baldwin, R.L. 1995. N- and C-capping preferences for all 20 amino acids in α -helical peptides. *Protein Sci.* **4**: 1325-1336.
- Doig, A.J., Chakrabarty, A., Klingler, T.M., and Baldwin, R.L. 1994. Determination of free energies of N-capping in α -helices by modification of the Lifson-Roig helix-coil theory to include N- and C- capping. *Biochemistry* **33**: 3396-3403.
- Facchiano, A.M., Colonna, G., and Ragone, R. 1998. Helix stabilizing factors and stabilization of thermophilic proteins: An X-ray based study. *Protein Eng.* **11**: 753-760.
- Finkelstein, A.V. 1977. Electrostatic interactions of charged groups in water environment and their influence on the polypeptide chain secondary structure formation. *Molek. Biol. (USSR)* **10**: 811-819.
- Finkelstein, A.V., Badretidinov, A.Y., and Pitsyn, O.B. 1991. Physical reasons for secondary structure stability: α -helices in short peptides. *Proteins* **10**: 287-299.
- Gill, S.C. and von Hippel, P.H. 1989. Calculation of protein extinction coefficients from amino acid sequence data. *Anal. Biochem.* **182**: 319-326.
- Harper, E.T. and Rose, G.D. 1993. Helix stop signals in proteins and peptides: The capping box. *Biochemistry* **32**: 7605-7609.
- Kiefhaber, T. and Baldwin, R.L. 1995. Intrinsic stability of individual α helices modulates structure and stability of the apomyoglobin molten globule form. *J. Mol. Biol.* **252**: 122-132.
- Kumar, S. and Bansal, M. 1998. Dissecting α -helices: Position-specific analysis of α -helices in globular proteins. *Proteins* **31**: 460-476.
- Lacroix, E., Viguera, A.R., and Serrano, L. 1998. Elucidating the folding problem of α -helices: Local motifs, long-range electrostatics, ionic-strength dependence and prediction of NMR parameters. *J. Mol. Biol.* **284**: 173-191.
- Lee, K.H., Xie, D., Freire, E., and Amzel, L.M. 1994. Estimation of changes in side chain configurational entropy in binding and folding: General methods and application to helix formation. *Proteins* **20**: 68-84.
- Mazur, A.K. and Abagyan, R.A. 1989. New methodology for computer-aided modelling of biomolecular structure and dynamics. 1. Non-cyclic structures. *J. Biomol. Struct. Dyn.* **6**: 815-832.
- Momany, F.A., McGuire, R.F., Burgess, A.W., and Scheraga, H.A. 1975. Energy parameters in polypeptides. VII. Geometric parameters, partial atomic charges, nonbonded interactions and intrinsic torsional potential for the naturally occurring amino acids. *J. Phys. Chem.* **79**: 2361-2381.
- Muñoz, V. and Serrano, L. 1994a. Elucidating the folding problem of helical peptides using empirical parameters. *Nat. Struct. Biol.* **1**: 399-409.
- . 1994b. Intrinsic secondary structure propensities of the amino acids, using statistical ϕ - ψ matrices: Comparison with experimental scales. *Proteins* **20**: 301-311.
- . 1995a. Elucidating the folding problem of helical peptides using empirical parameters. II. Helix macrodipole effects and rational modification of the helical content of natural peptides. *J. Mol. Biol.* **245**: 275-296.
- . 1995b. Elucidating the folding problem of helical peptides using empirical parameters. III. Temperature and pH dependence. *J. Mol. Biol.* **245**: 297-308.
- . 1997. Development of the multiple sequence approximation within the

- AGADIR model of α -helix formation: Comparison with Zimm-Bragg and Lifson-Roig formalisms. *Biopolymers* **41**: 495–509.
- Nemethy, G., Pottle, M.S., and Scheraga, H.A. 1983. Energy parameters in polypeptides. 9. Updating of geometrical parameters, nonbonded interactions and hydrogen bond interactions for the naturally occurring amino acids. *J. Phys. Chem.* **87**: 1883–1887.
- Petukhov, M., Yumoto, N., Murase, S., Onmura, R., and Yoshikawa, S. 1996. Factors that affect the stabilization of α -helices in short peptides by a capping box. *Biochemistry* **35**: 387–397.
- Petukhov, M., Kil, Y., Kuramitsu, S., and Lanzov, V. 1997. Insights into thermal resistance of proteins from the intrinsic stability of their α -helices. *Proteins* **29**: 309–320.
- Petukhov, M., Muñoz, V., Yumoto, N., Yoshikawa, S., and Serrano, L. 1998. Position dependence of non-polar amino acid intrinsic helical propensities. *J. Mol. Biol.* **278**: 279–289.
- Petukhov, M., Uegaki, K., Yumoto, N., Yoshikawa, S., and Serrano, L. 1999. Position dependence of amino acid intrinsic helical propensities II: Non-charged polar residues: Ser, Thr, Asn, and Gln. *Protein Sci.* **8**: 2144–2150.
- Prieto, J. and Serrano, L. 1997. C-capping and Helix stability: The Pro C-capping motif. *J. Mol. Biol.* **274**: 276–288.
- Qian, H. 1993. Single-residue substitution in homopolypeptides: Perturbative helix-coil theory at a single site. *Biopolymers* **33**: 1605–1616.
- Richardson, J.S. and Richardson, D.C. 1988. Amino acid preferences for specific locations at the ends of α helices. *Science* **240**: 1648–1652.
- Richmond, T.J. 1984. Solvent accessible surface area and excluded volume in proteins. Analytical equations for overlapping spheres and implications for the hydrophobic effect. *J. Mol. Biol.* **178**: 63–89.
- Stapley, B.J. and Doig, A.J. 1997. Free energies of amino acid side-chain rotamers in α -helices, β -sheets and α -helix N-caps. *J. Mol. Biol.* **272**: 456–464.
- Stapley, B.J., Rohl, C.A., and Doig, A.J. 1995. Addition of side chain interactions to modified Lifson-Roig helix-coil theory: Application to energetics of phenylalanine-methionine interactions. *Protein Sci.* **4**: 2383–2391.
- Street, A.G. and Mayo, S.L. 1999. Intrinsic β -sheet propensities result from van der Waals interactions between side chains and the local backbone. *Proc. Natl. Acad. Sci.* **96**: 9074–9076.
- Strehlow, K.G. and Baldwin, R.L. 1989. Effect of the substitution Ala—Gly at each of five residue positions in the C-peptide helix. *Biochemistry* **28**: 2130–2133.
- Strop, P., Marinescu, A.M., and Mayo, S.L. 2000. Structure of a protein G helix variant suggests the importance of helix propensity and helix dipole interactions in protein design. *Protein Sci.* **9**: 1391–1394.
- Sun, J.K., Penel, S., and Doig, A.J. 2000. Determination of α -helix N1 energies after addition of N1, N2, and N3 preferences to helix/coil theory. *Protein Sci.* **9**: 750–754.
- Swindells, M.B., MacArthur, M.W., and Thornton, J.M. 1995. Intrinsic ϕ , ψ propensities of amino acids, derived from the coil regions of known structures. *Nat. Struct. Biol.* **2**: 596–603.
- Taddei, N., Chiti, F., Fiaschi, T., Bucciantini, M., Capanni, C., Stefani, M., Serrano, L., Dobson, C.M., and Ramponi, G. 2000. Stabilisation of α -helices by site-directed mutagenesis reveals the importance of secondary structure in the transition state for acylphosphatase folding. *J. Mol. Biol.* **300**: 633–647.
- Thomas, S.T., Loladze, V.V., and Makhatadze, G.I. 2001. Hydration of the peptide backbone largely defines the thermodynamic propensity scale of residues at the C' position of the C-capping box of α -helices. *Proc. Nat. Acad. Sci.* **98**: 10670–10675.
- Villegas, V., Viguera, A.R., Aviles, F.X., and Serrano, L. 1995. Stabilization of proteins by rational design of α -helix stability using helix/coil transition theory. *Fold. Des.* **1**: 29–34.
- Viguera, A.R. and Serrano, L. 1995. Experimental analysis of the Schellman motif. *J. Mol. Biol.* **251**: 150–160.
- Vriend, G. 1990. WHAT IF: A molecular modeling and drug design program. *J. Mol. Graph.* **8**: 52–56.
- Warshel, A. and Papazyan, A. 1998. Electrostatic effects in macromolecules: Fundamental concepts and practical modeling. *Curr. Opin. Struct. Biol.* **8**: 211–217.
- Wesson, L. and Eisenberg, D. 1992. Atomic solvation parameters applied to molecular dynamics of proteins in solution. *Protein Sci.* **1**: 227–235.
- Zimmerman, S.S., Pottle, M.S., Nemethy, G., and Scheraga, H.A. 1977. Conformational analysis of the 20 naturally occurring amino acid residues using ECEPP. *Macromolecules* **10**: 1–9.